图书情报工作
第64卷 第4期 2020年2月

新兴技术识别方法研究综述

■ 王玏[1,2,3] 吴新年[1,2,3]

[1] 中国科学院西北生态环境资源研究院 兰州 730000 [2] 中国科学院兰州文献情报中心 兰州 730000

[3] 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100049

摘 要：[目的/意义]对国内外新兴技术识别方法进行系统梳理，总结研究现状，分析存在问题，为今后新兴技术识别方法的研究提供参考。[方法/过程]在文献调研与计量分析的基础上，归纳总结目前主流的新兴技术识别方法，结合典型案例分析各类方法的特征及其在新兴技术识别过程中的优缺点，并为未来新兴技术识别方法研究提出建议。[结果/结论]目前的新兴技术识别方法既有定量的，也有定性的，常见的方法包括基于测度模型、文献计量以及文本挖掘的方法，且越来越倾向于多方法融合应用，不过现有研究仍然存在一些不足。未来对新兴技术识别方法的研究应该加强对新兴技术本质特征的探究，加强具体识别方法对新兴技术的解释意义，以及从数据源、识别方法两方面共同提升新兴技术识别的时效性；同时，还应该加强对新兴技术识别中多源数据有效融合与应用方法的探索。

关键词：新兴技术 技术识别 识别方法

分类号：G250

DOI：10.13266/j.issn.0252-3116.2020.04.014

随着技术演变的速度和复杂性不断提升以及技术之间的交叉性、相互依赖性的增强，从众多新技术中快速识别出具有较大发展潜力的新兴技术的难度也在不断提升，因此近年来关于新兴技术的识别预测成为相关研究领域日益受关注的研究主题。

关于新兴技术的定义及其特征学界一直没有达成统一，目前最广泛认可的定义是宾夕法尼亚大学沃顿商学院在2000年出版的著作 *Wharton on Managing Emerging Technologies* 中给出的定义——新兴技术是建立在科学基础上的创新性技术，它们可能创立一个新行业或改变某个老行业[1]。关于新兴技术的特征，最广泛认可的是 D. Rotolo 等[2]提出的新颖性、相对快速的增长、一致性、突出的影响以及不确定性。而对于新兴技术的识别，学者们在具体操作中通常主要考虑新兴技术的“新”与“兴”，即新颖性和相对快速的发展。

从国家层面来讲，加强新兴技术识别有助于整体科研布局优化以及做好长期规划；从企业层面来讲，加强新兴技术识别有助于确定研发重点、投资方向以及降低投资风险。同时，新兴技术识别是研究前沿技术、突破性技术、颠覆性技术等的基础与前提，一些新兴技术可能就是前沿技术，也可能发展成为突破性技术或颠覆性技术。加强新兴技术的识别与预测方法研究，不断提升对各领域新兴技术识别与预测的准确性和效率，具有十分重要的战略与现实意义。正是基于对此问题的深刻理解，笔者拟在对国内外当前常用的新兴技术识别方法进行深入系统调研的基础上，对这些常见识别方法进行分类分析，以期为后续研究提供参考。

为了综观新兴技术识别预测方法研究总体态势并基于核心论文进行分类分析，笔者选取 Web of Science 和中国知网作为数据源。Web of Science 检索策略为：TS =((identif* or forecast* or discover* or detect* or predict* or assess* or evaluat* or anticipat* or demystif* or measur*) and ("emerg* technolog*" or "technolog* emerg*" or "emerg* topic*"or "technolog* potential")),索引=(SCI-EXPANDED, SSCI, CPCI-S),时间跨度=所有年份，共检索到相关文献9 608篇。中国知网检索策略为：SU=(‘新兴技术’+‘新兴主题’+‘技术潜力’) AND SU=(‘识别’+‘预测’+‘测度’+‘发现’+‘评估’+‘评价’+‘探测’),来源类别=全部期刊，发表时间=不限，共检索

作者简介：王玏(ORCID:0000-0001-9065-8495)，硕士研究生；吴新年(ORCID:0000-0002-7865-9548)，研究员，博士，博士生导师，通讯作者，E-mail:wuxn@lzb.ac.cn。

收稿日期：2019-07-18 修回日期：2019-09-11 本文起止页码：125-135 本文责任编辑：徐健

到相关论文 203 篇(检索时间:2019 年 7 月 16 日)。通过对检索到的文献进行基本指标的计量分析,了解了该主题领域研究截止目前的基本态势,在此基础上基于一些影响力指标的分析和人工判读,以及对重要文献的参考文献的补充追踪,选取能够表征该领域研究历程且影响力较高的论文,以及相关度高且能够反映最新研究进展的论文作为核心文献,共计 55 篇(其中英文文献 38 篇、中文文献 17 篇)。笔者针对新兴技术识别方法的分类分析主要基于这些核心文献展开。

1 新兴技术识别方法研究总体态势

国内外关于新兴技术识别预测方法的研究从 20 世纪 80 年代就已开始,此后年度论文产出量、参与的研究机构数量及研究人员数量在 1995 年后开始逐渐上升,2010 年后进入快速增长期。从涉及的学科领域看,国内发表相关论文最多的学科是工商管理、数量经济学、图书情报档案学以及科学学与科技管理,国外文献的学科分布则相对广泛,涉及计算机科学、管理科学、商学以及其他部分自然科学领域。发文较多的主要期刊包括 *Technological Forecasting And Social Change*、*Scientometrics*、《情报杂志》《情报理论与实践》以及《情报学报》等。著名的学者包括 A. L. Porter、P. Shapira、J. Youtie、黄鲁成、吴菲菲、周源等。如图 1 所示:

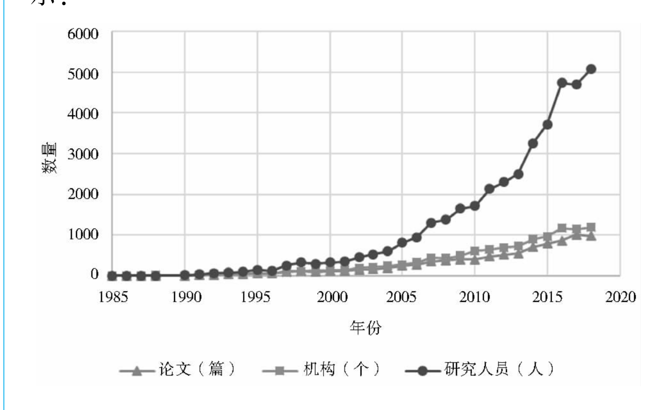


图 1 全球新兴技术识别预测领域研究发展趋势

新兴技术识别预测方法研究涉及的核心关键词包括:文献计量、专利分析、机器学习、数据挖掘、文本挖掘、大数据、引文网络、社会网络分析、内容分析、主题探测、模型、技术路线图、聚类、分类等。如图 2 所示:

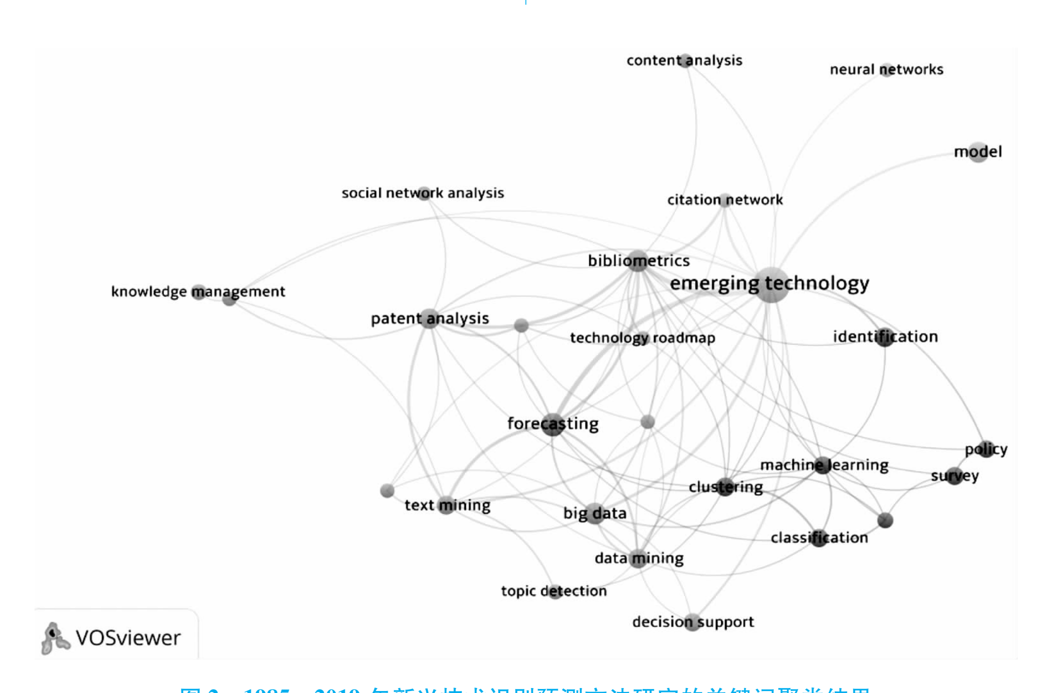


图 2 1985 – 2019 年新兴技术识别预测方法研究的关键词聚类结果

新兴技术识别预测方法研究关键词年度分布如表 1 所示。在新兴技术识别预测中,文献计量方法的应用比例最高,机器学习近年来发展最为迅速,近 5 年深度学习方法开始越来越多地应用于新兴技术识别预测。

从以上分析中可以看出,近年来新兴技术识别预测方法的研究热度不断上升,且受到较多领域的广泛关注;同时,机器学习、文本挖掘等方法在其中的应用近几年发展尤为迅速。而新兴技术识别是整个新兴技术预测领域中最为关键的部分,只有实现对新兴技术的早期识别,才能进行后续的对新兴技术未来发展目标、可能途径及资源条件等的预测及评估。

表 1　新兴技术识别预测方法研究关键词年度分布

关键词＼年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
文献计量											2	4	6	4	8	6	10	8	13	14	12	17	19
专利分析											1	1	3	3	4	2	1	3	7	11	8	14	9
网络分析								1				1	4		2	6	4	7	4	9	9	13	13
内容分析							1		1	1			1		2	4	2	2	3	7	2	10	6
数据挖掘	1	1	1	3	2	1		1	4	1		2	2	3	4	6	6	8	8	10	12	18	11
文本挖掘					1		1	1			2	3	1	1		3	6	2	4	6	5	15	8
机器学习										2		5		2	1		2		7	8	10	20	30
深度学习																				1	2	9	12

2　常见的新兴技术识别方法分析

从广义概念来讲,新兴技术识别是技术预测的一个子类。技术预测是对未来某个时间特定技术本身的产生、特征和影响进行的预先测定[3];而“识别”是基于现有事实进行的分类或定性,“新兴技术识别”则是基于现有相关数据中的信号判定哪些技术是新兴技术。技术预见是对技术发展及其与社会和环境相互作用的长期未来前景进行描绘的系统过程[3]。

从相关研究文献梳理中可以发现,德尔菲法、情景分析法、技术路线图法、文献计量法、测度模型法等是技术预测领域较常用的方法,其中,德尔菲法、情景分析法、技术路线图法等更多见于技术预见活动(如日本政府自 20 世纪 70 年代以来每 5 年一次的大型技术预见活动),而这些方法若用于新兴技术识别,则常常需要与文献计量、文本挖掘等定量方法相结合。

近年来,新兴技术识别领域发展较快的主要是基于测度模型、文献计量、文本挖掘的分析方法以及这些方法的融合应用,而且各类方法似乎各有优势,当然也分别存在一些需要改进的方面。

2.1　基于测度模型的新兴技术识别方法

许多学者会考虑新兴技术的识别标准,从而选取相关指标,构建相应的测度模型对新兴技术进行识别。基于测度模型的新兴技术识别方法的出现对整个新兴技术识别方法论体系的完善产生了重要影响,说明新兴技术识别方法开始逐渐考虑新兴技术的内核,并将识别流程模式化以更便捷地应用于新兴技术识别。从本质上来讲,基于一定的指标体系构建模型进行新兴技术识别的方法在科学性方面已经有了很大的改进,通过不断改进指标的赋值和计算办法可以促进提高新兴技术识别的准确性和效率,近年来许多学者都积极投入这方面研究并取得了很好的进展。例如,黄鲁成等[4]从新兴技术特征出发,在属性集和属性测度理论基础上提出了属性综合评价和决策系统,构建了新兴技术识别指标体系模型,包括技术指标和市场指标两类指标,其中技术类指标又包括技术先进性、技术独特性、技术复杂性、研发投入、研发产出、形成技术标准可靠性 6 个二级指标,市场类指标又包括市场规模－前景、市场增长速度、改变市场格局的可能性和为客户创造价值 4 个二级指标,并利用构建的模型从 10 项新材料技术中识别出 1 项新兴技术、4 项新技术、5 项一般技术;该方法有效地克服了以往方法中专家打分的主观随意性, 保证了新兴技术识别的科学性、客观性和准确性,但是有些指标的可计算性和取值办法的科学合理性还有待商榷。任智军等[5]融合频度、年数、归属度等内容性指标以及增长率、相对增长率、作者占有率等趋势性指标构建特征向量,将 LDA 主题模型识别的技术主题与 Gartner 的新兴技术主题通过特征向量计算相似度,最终从 1 000 个技术主题中识别出了 25 个相似度最高的技术主题作为新兴技术。类似的,还有 M. N. Kyebambe 等[6]的研究,其认为新兴技术的踪迹可以在其完全出现前几年在专利数据库中找到,因此研究了过去某个特定时间点后催生了一项新兴技术的一系列专利所具有的特性,包括专利权项数、引用次数、对非专利文献的引用次数、技术生命周期、被引技术相似度以及被引专利受让人相似度,并以这些特征为每个专利创建专利特征向量并构建模型,通过聚类形成不同技术集群,训练模型将专利集群标记为新兴技术或非新兴技术;但是,由于在该研究中仅使用 7 个特性来区分技术集群,使得技术集群类别不够细化,包

含技术范围较广，即模型识别新兴技术的敏感度不够高，因此未来可能需要增加更多的有效特征因子以提高模型的准确度。K. Song 等[7]首先从专利耦合网络中通过专利内容相似性进行聚类，找出离群专利，并从中识别出最近发表的且与现有专利几乎没有相似之处的专利作为候选新兴技术，之后通过专利权范围、专利应用范围、专利发明者数量等技术特征指标以及市场需求等市场特征指标构建模型识别出真正具有发展潜力的新兴技术；但是，由于该研究中提出的市场特征指标是基于客户评价数据进行分析，因此只适用于客户评价数量积累到足够提供可靠的市场需求信息的领域，具有一定的局限性。Q. Wang 等[8]认为新兴技术识别指标体系应该与新兴技术定义高度吻合，因此他基于新兴技术特征提出了新兴技术主题的识别标准——新颖性、相对快速的增长、科学影响力和一致性，并将其通过文献计量指标量化后构建模型用以识别新兴技术。

总体而言，基于特定测度模型的新兴技术识别方法是近几年发展起来并且应用较广泛的方法，此类方法的主要问题在于指标选取的科学性及合理性难以界定，各个指标的表征力判定、权重确定以及模型构建方面都还有许多值得探讨的问题需要解决，另外有些指标的取值办法和计算方法也是现实中常常令人头痛的问题。而且，由于目前新兴技术的定义及其特征并没有统一的结论，因此基于其内核特征及定义的模型构建也较容易引起争议。

2.2 基于文献计量学的新兴技术识别方法

近年来，随着各类文献计量软件的快速发展，基于文献计量学的方法在新兴技术识别方面的应用越来越广泛。基础研究是创新的源头，科学文献中往往蕴含着丰富的技术创新信息，基于科学文献（包括论文、专利等）在一定程度上可以探测出人类技术创新与发展的脉络。正是在这种思想指导下，基于文献计量学的方法在新兴技术识别中日益得到重视和应用。

2.2.1 基于科学论文数据的新兴技术识别方法

科学论文数据用于新兴技术识别可以在技术开发周期的早期捕获其动向信息，但也正因为其还在基础研究阶段，单独使用科学论文数据很难识别出真正意义上的“新兴技术”，而更多的是识别新兴主题和新兴研究领域。但是，在基于技术全生命周期的识别预测中其可以作为针对相关技术主题基础研究阶段发展态势的分析，并可以作为对相关技术研发是否可以获得基础研究领域更持久的创新支持和新知识供给的考察，亦即研判该技术主题持续创新的潜力与发展趋势。因此，基于科学论文的计量分析对于发现与识别新兴技术也是很重要的一种方式。

首先，不得不提的是传统的文献计量指标在基于科学论文数据的新兴技术识别中的应用。例如，C. Mund 等[9]选取了 5 个不同的领域，通过期刊规模、期刊年龄、参考文献年龄、作者数量、作者合作情况等文献计量指标计算识别各领域的新兴技术主题，并对这些指标在不同领域识别新兴技术主题的效果差异进行了对比分析，结果表明：不同指标对不同领域的新兴技术识别效果差异较为显著。比如，“研究团队规模”指标在识别医学领域新兴技术时具有更高的辨识度，而“期刊规模”和“期刊年龄”在识别工程领域新兴技术时具有更高的辨识度。

除传统的文献计量指标以外，有一些学者也会设置一些创新性的计量指标并结合其他方法识别新兴技术。例如，E. Schiebel 等[10]基于文献过滤指标（高扩散度、低扩散度、相对频率等）筛选光电子器件领域各研究阶段重要术语，构建出了领域术语演化的扩散模型，并结合历时聚类分析（diachronic cluster analysis）来识别领域新兴技术主题。

基于科学论文数据的新兴技术识别通常还会采用论文的直接引用网络、共引网络、耦合网络等等中的一种或几种进行聚类分析。例如，K. Fujita 等[11]在 3 种引文网络（直引、共引、耦合网络）中对文献聚类后，通过计算簇团文献规模、簇团中文献的平均出版年份、簇团中文献的文本相似性及簇团的密度，评估 3 种加权引文网络在不同领域识别新兴技术的效果，并结合氮化镓、纳米碳两个技术领域进行了实证研究，发现加权引文网络比非加权引文网络在检测新兴研究前沿方面更有用，并且以引用频次作为权重检测新兴研究前沿比以发表年份差异、文献相似性、关键词相似性等作为权重更为有效。H. Small[12]基于直接引用和共同引用网络，结合使用差异函数遴选出新兴技术主题，通过搜索与该主题或其主要研究人员相关的近期主要成果来评估新兴技术主题，生成新兴技术主题清单供决策者决策。在新兴技术识别中使用最为广泛的引文网络就是共引网络[13]，但是共引网络生成时以论文共被引为依据，因此可能会忽略一些最新的文献，这会在一定程度上影响新兴技术识别的有效性和准确性。

一些学者为了有效减少论文时滞性对新兴技术识

别的影响,因此选择会议论文作为新兴技术识别的数据源,其中包括 T. Furukawa[14]、王燕鹏[15]等学者。而在某些学科领域(如计算机科学等),会议论文本来就是最为重要的数据源,应该作为新兴技术识别的首选数据源。王燕鹏等[15]将关键技术分为热点技术、共性技术以及新兴技术,并选择人工智能领域高水平国际会议论文作为数据源,通过 Sci2 tool 工具的 Burst Detection 算法检测突现词,识别出人工智能领域的 15 项新兴技术。J. Kleinberg[16]于 2002 年提出的 Burst Detection 算法在新兴技术识别中有较为广泛的运用,该算法被嵌入许多计量工具,如 Citespace、Sci2 以及 Workbench 等。C. M. Chen[17]、H. Guo[18]等学者也曾采用关键词突现检测的方法识别新兴科技主题,但是仅仅依靠突现检测识别新兴技术可能缺乏准确性及有效性,应当适当结合其他指标以及权重设置等进行识别预测。

总体来说,基于科学论文数据的文献计量方法在新兴技术识别中应用较为广泛,但通常以“新兴技术主题识别”的方式呈现,这是因为将科学论文作为新兴技术识别唯一的数据来源时其核心信息即为研究主题,其识别结果本质上是未考虑技术性能与市场的切合度以及未来应用前景还不明朗的新兴技术,这也是仅使用科学论文作为数据源识别新兴技术的一个局限性。并且由于传统的文献计量学涉及到的计量指标有限,因此应用于新兴技术识别时需要与其他方法相结合并且在具体指标上要有更进一步的个性化的创新,如 E. Schiebel 等[10]提出的用于构建术语扩散模型的扩散度等指标。

2.2.2 基于专利数据的新兴技术识别方法

A. Pilkington[19]早在 2004 年就开始探索专利作为描绘技术发展的信息来源的潜力,认为专利在预测技术商业化前景及新兴技术方面的价值有待开发。这是因为专利是技术信息的载体,相比于论文数据更能反映技术层面的发展,由此基于专利数据的文献计量方法也成为文献计量学在新兴技术识别的应用中最主要的部分,相关文献较多,且近年来数量上升极快。

基于专利数据进行新兴技术识别常会将专利引用次数作为专利价值的衡量指标。但是,目前很多学者都在质疑专利被引次数是否能真正体现专利未来价值及影响力[20]。兰德公司研究报告 *Identification and Analysis of Technology Emergence Using Patent Classification* 中也明确指出专利的引用与其价值之间的函数关系并非线性,并且二者之间的正相关性也没有曾经认为的那么理想,该研究以 USPTO 的专利分类系统和相关数据为基础,通过对纳米领域专利申请数量以及领域相关技术分类数量之间的关系随时间的变化进行分析,并结合 S 曲线识别出了量子点技术为该领域的新兴技术[21]。

基于专利数据的新兴技术识别也常采用专利的直引、共引网络等,因此也常与社会网络分析方法相结合。例如,P. C. Lee 等[22]在专利直引网络中运用社会网络分析法,计算度中心性、中介中心性、接近中心性用于表示几种技术演化机制并揭示新兴技术;但是,该研究中展示的专利引用图没有考虑时间跨度,如果作者能够建立二维专利引用图,则可以观察领域技术专利随时间的发展情况,这样就能够更好地识别新兴技术。T. S. Cho 等[23]也将专利引用网络与社会网络分析相结合,并用于识别台湾技术创新体系中的核心技术和新兴技术,其中对新兴技术的识别主要采用结构洞分析方法,识别出了 5 项跨越不同专利子类组别并有机会与其他专利类别融合在一起的新兴技术。李瑞茜[24]以中国大陆在 35 个技术领域获得中国授权的有效发明专利共类数据为样本,构建了非对称的技术知识流网络,并采用社会网络分析的中心度、结构洞和中间人分析方法,探索了技术关联结构特征,识别了技术关联网络的核心技术、中介技术和新兴技术。黄璐等[25]利用德温特专利数据,基于加权共词网络进行链路预测,综合考虑网络局部、路径以及随机游走这三类网络结构信息,构建了用于预测技术发展的动态网络,并围绕新颖性与影响力两个维度识别新兴技术,最后应用钙钛矿材料领域相关数据进行实证分析,部分结论与当前学术界的共识基本保持一致,证明了该方法的有效性。到目前为止,学者们将社会网络分析法用于新兴技术识别中时所采用的方法和指标重合度较高,说明社会网络分析方法在新兴技术识别中的应用较为成熟,但是就不同指标在新兴技术识别中所代表的内涵及其与新兴技术的关联还有待进一步阐释。

在新兴技术识别中,对专利信息的可视化也不仅仅局限于计量软件对引用网络等的展示,学者们将专利中的新兴主题的可视化进行拓展,形成各类专利地图。例如,S. Lee 等[26]提出了一种基于关键词创建专利地图的方法,首先确定专利“空缺”,即地图中专利密度较小但规模较大的空白区域,然后对专利“空缺”的有效性进行测试,即计算“空缺”中专利申请的增长

率,以区分其为新兴技术还是正在衰落的技术。但是,目前的专利信息可视化较少能做到二维空间的映射,这将是未来的一个研究方向,以便于更好地观察技术随时间的发展变化,以此确定技术的发展潜力。

也有许多学者不断对专利指标进行完善和拓展,以提升对新兴技术的识别效果。例如,G. De Rassenfosse 等[27]为了解决专利数据的覆盖面问题,提出了一个新的专利指标以识别新兴技术,该指标对一个国家发明人提交的所有优先专利申请进行计数,而不论提交的专利局是哪一个。P. Erdi 等[28]定义了一种称为引文向量的预测因子来表征技术发展,引用向量的每个坐标与该专利在特定时间内被特定技术类别中的其他专利所引用的相对频率成比例,并以引文向量度量专利之间的相似性,在专利共引网络中以专利相似性聚类得出新兴技术集群,并观察集群结构随时间的动态变化以识别新兴技术领域;但是,该方法对专利引用网络的简化以及对引用向量权重设置的合理性还有待进一步验证。李蓓等[29]提出了一种基于专利引用耦合聚类的新兴技术识别方法,并选取纳米技术领域作为研究对象,通过计算技术集群的专利授权时间、专利数量年增长率、专利权利要求项数 3 个指标识别出了纳米领域的 2 项新兴技术;但是,该方法中只是对耦合连接数大于等于 10 的核心专利展开了聚类分析,而新兴技术刚刚萌芽时,可能与已存在的技术间联系会非常少,因此,该方法很可能会忽略掉一些尚处于萌芽期的新兴前沿技术信息。G. Kim 等[30]也提出了类似的用以识别新兴技术的专利指标化计算方法。

然而,相比于科学论文数据,专利数据的时滞性问题更为严重。为了弱化这一问题的影响,一些学者探索性地提出了运用专利数据时提高新兴技术识别时效性的方法。例如,G. De Rassenfosse 等[27]提出了一个用以识别新兴技术的指标,通过对一个国家发明人提交的所有优先专利申请进行计数,该指标捕捉的是申请的专利,而不是授予的专利,因此在一定程度上保证了专利数据的及时性;但是,这里又会引发另外一个问题,即还未获得授权的专利质量可能又难以保证。A. Breitzman 等[13]则提出基于 Emerging Clusters 模型来近乎实时地而非回顾性地定位新兴技术,首先确定给定时间段内的热门专利,然后识别出引用热门专利的下一代专利并进行聚类形成专利集群,再通过专利集群的 4 个特征(公共部门比例、科学指数、原创指数、参考指数)对其进行评分和排名,其中得分最高的被定义为"新兴集群",以此识别新兴技术;但是,该研究中对下一代专利的评分机制包含的特征指标较为局限,仅涉及到专利元数据,今后的研究也许可增加一些内容分析的指标,如将数据范围进一步扩展至专利摘要等的挖掘分析,可能有利于提升方法的准确性。C. Lee 等[20]也提出了一种弱化专利数据时滞性的方法,首先从美国专利商标局数据库中提取出 18 个技术特征指标作为输入指标,这部分指标可在相关专利发布后立即获取数据,包括技术创新级别、技术范围、总体技术实力等,然后再提取出 3 个潜在影响力指标作为输出指标,包括未来 3 年、5 年、10 年内的被引次数,通过多层前馈神经网络捕获输入指标和输出指标之间的非线性关系,以该非线性关系预测当前专利的未来价值,并结合两个定量指标"emergingness"和"trend"识别新兴技术。但是,该方法也还存在一些明显的问题:①其仅仅采用了专利被引次数来体现专利价值,显得有些单薄,如果增加更多样的输出指标(如专利授权量、专利转让价格等)加强技术特性与未来影响力之间的关系计算与研判,效果可能会更好一些;②其通过神经网络计算出的技术特性与未来影响力之间的非线性关系缺乏可解释性,应该对此给出科学、合理的论证。

总体而言,专利作为技术信息的重要载体,能更好地体现技术创新层面的发展动向,但是由于专利审核周期较长等原因,其用于新兴技术识别的时滞性也较为严重,因而专利作为新兴技术识别中应用最广泛的数据源,在今后的研究中如何尽可能地降低其时滞性对预测结果的影响将是一个值得关注的问题。

2.2.3 基于论文与专利数据融合计算的新兴技术识别方法

H. M. Jarvenpaa 等[31]认为关于技术预测的最常见的批评之一就是使用单一数据库作为数据来源,应该使用不同来源的数据以克服单一来源数据可能导致的偏见。而新兴技术识别领域的大多数研究都选择了单一数据源(专利数据或论文数据),仅有个别学者将论文、专利数据结合起来进行分析。例如,王凌燕等[32]尝试建立了一套新兴技术主题判定评价指标体系,包括 3 类文献计量指标(主题词词频的变化指标、战略坐标图指标和共词网络指标)和 4 个专利特征指标(专利申请量增长率、技术生长率、专利家族平均成员数变化、专利他引率),通过文本聚类技术、共词战略坐标分析、共词网络分析、专利分析等共同识别出了工业生物领域的新兴技术,并通过咨询相关领域专家、对

比国家相关政策及相关技术项目资金投入情况对识别结果的有效性进行了检验。同样地,T. Ogawa 等[33]首先通过聚类燃料电池领域学术论文的引文网络来确定子研究领域,然后通过测量和分析子研究领域的专利相关性和时间序列变化以识别该领域的新兴技术,并选择固体高分子型燃料电池(PEFC)进行了验证,结果表明,在一个论文数量较少、平均出版年份较近、专利相关性较低的技术领域,其随后几年很可能会有较大的发展潜力。

越来越多的研究表明,将论文与专利数据相结合进行分析,虽然可能会在一定程度上增加分析工作量,但由于不同数据源之间可以互为补充,有利于减少因为使用单一数据源所带来的分析误差。进一步地,也有学者提出可以尝试增加更多的数据源(如会议信息、项目数据、商业报道数据等),通过多源数据的融合计算进行新兴技术的识别,这种新型方法也许能够带来更好的识别效果。但是因为多源数据的多层次、多类别,对其进行有效融合难度较大[34];同时,对分析数据集中各数据源数据的比例设置问题也缺乏标准,仅仅是对论文与专利数据的融合方面尚且缺乏深入探讨[35],部分学者在基于多源数据融合计算识别新兴技术方面进行了一些实践性探索,但对其中的机理研究还远远不够。徐路路等[36]认为进行多数据源融合的前提是拟融合数据源的相关性分析,因此假设拟融合数据源要满足文本特征要素相关、发展趋势相关、主题相关以及情景相关,作者选取论文数据、基金项目数据、专利数据 3 种数据源进行相关性分析,并基于 PLDA 主题模型对多源文本的摘要信息进行融合计算,识别出了石墨烯领域的 4 个新兴技术主题;但是,该研究对多源数据的融合力度还有所不足,仅仅是简单融合了多源文本的摘要数据,未来还可以研究针对不同数据源的评价指标体系,并通过参数设置真正实现多源文本主题信息的深度挖掘与融合计算。

2.3 基于文本挖掘分析的新兴技术识别方法

随着文本挖掘技术的发展,将文本挖掘分析方法应用于新兴技术识别也是目前日益热门的研究点之一。目前其在新兴技术识别中的应用主要是挖掘科学论文或专利的语义信息,最常用的两种方法分别是基于主题模型的方法和基于结构语义分析的方法。基于主题模型的方法主要用以对领域技术主题及相关术语进行分类,基于结构语义的方法主要用以定义及提取新兴技术关键词。总之,文本挖掘分析方法应用于新兴技术识别是对新兴技术识别流程与技术方法的重大改进,能够高通量地、高效地处理海量科学文献及专利数据,减少人工筛选、处理数据的繁琐。

2.3.1 基于主题模型的方法

主题模型是以非监督学习的方式对文档的隐含语义结构进行聚类统计[37]。新兴技术识别中最常见的主题模型就是 LDA 主题模型,其将每一篇文档表示为一些主题所构成的一个概率分布,而每一个主题又被表示为很多词所构成的一个概率分布,以此构建文档与技术主题之间的关系,从而识别领域内技术主题并进行分类。例如,任智军等[5]采用 LDA 模型获取技术主题,并通过特征向量与 Gartner 的新兴技术主题进行相似度计算,通过设定相似度阈值,遴选出有限个技术主题交给专家进行判定,该方法减少了专家的主观性和待分析新兴技术的数量,提高了专家的工作效率。董放等[38]则通过 LDA 主题模型和 SVM(支持向量机)分类模型共同处理论文摘要数据,实现对论文按照技术主题的分类,并用 ARIMA 模型预测技术领域未来论文数量的变化趋势,从而识别技术领域发展潜力。D. Choi 等[39]基于 LDA 主题模型确定领域主题及相关术语,之后通过计算每个主题当前的专利份额以及该主题专利份额随时间的变化率,识别出物流领域 4 种不同的技术主题(主导主题、新兴主题、饱和主题和衰退主题)。

因为 LDA 主题模型是典型的无监督机器学习,结果可控性较低[38],因此有学者开始使用有监督机器学习方法,以提高主题模型对技术分类的准确性。有监督学习与无监督学习的本质差别在于有监督学习必须有标签。例如,周源等[40]提出了一种基于机器学习主题模型的新兴技术识别方法,将专家组的领域知识与判断融入机器学习过程中,从而提高了机器学习主题模型的准确度与识别新兴技术的能力。其后,周源等[41]在此基础上进一步完善了该方法,其基于半监督主题聚类模型,从词级技术主题语义描述上升至句级技术主题语义描述,即以句子的形式描述技术主题聚类结果,进一步加强了技术语义解释能力。

技术主题的筛选与分类更进一步则是技术术语的筛选与分类,是对技术主题的一致化、规范化处理,是作为新兴技术识别预测前期工作的进一步深化。刘宇飞等[42]认为专利文献是技术情报最为重要的来源,但是由于新兴技术缺乏术语标签,对其挖掘及抽取难度较大,因此引入深度迁移学习的思想,首先抽取专利摘

要，通过引入专家知识构建标签语库，然后提取标签特征，通过 Bi-LSTM 模型实现源数据特征的迁移学习，最后针对输出结果，通过聚类分析，结合专家判断，将技术术语划分技术类别，该模型有效实现了技术术语自动识别并过滤高频非术语词串，且其恰当结合了专家知识，进一步保证了识别结果的有效性，其技术分类结果也有利于研究人员对新兴技术做进一步的识别及预测，这也是深度学习在新兴技术识别中运用的典型案例。

总之，基于主题模型的机器学习方法并非真正用于“识别”新兴技术，而是对新兴技术相关主题及术语进行前期处理，包括筛选与分类，其极大地降低了新兴技术识别中的人工成本。但是，“新兴技术”由于其新颖性和模糊性的特征，完全依赖算法的识别方法也显得说服力不足，因此，还应该适当结合专家意见，做到主客观相结合，以提高新兴技术识别的准确性。

2.3.2 基于结构语义分析的方法

因为新兴技术术语还不具备完全的一致性与规范性，因此有效定义及提取新兴技术领域的关键词是一项困难的任务，甚至对专家来说也是如此[20,43]。而基于结构语义的方法应用于新兴技术识别时，最大的优点在于能够自动提取专利的属性、功能、关联关系等，在一定程度上消除或弱化了上述问题。

基于结构语义的分析方法在新兴技术识别中的应用，最常见的就是 SAO 结构语义分析法。例如，李欣等[44]针对目前专利引用信息的滞后性和关键词聚类既不能深入反映专利文献表达的主题，也不能很好地揭示技术主题之间的关联关系等问题，提出了基于 SAO 结构语义分析法的新兴技术识别模型，首先抽取专利权利要求项中的 SAO 结构，再利用改进的语义相似度算法对专利文本进行聚类，最后结合基于时间切片的专利地图识别新兴技术，并以钙钛矿太阳能电池技术为例进行了实证研究，验证了该模型的可行性和有效性。随后，C. Yang 等[45]又提出了一种改进的 SAO 网络，该网络可计算节点之间的关系强度，并结合社会网络分析方法，基于结构洞、节点度数分布变化、网络中心度变化等进行特定领域新兴技术识别及其发展趋势的预测，并以石墨烯技术领域进行了实证研究，验证了其方法的有效性。

也有学者将 SAO 结构语义分析法与 TRIZ 理论相结合识别新兴技术。例如，翟东升等[46]在 TRIZ 理论基础上，通过 SAO-C 结构语义分析法抽取专利语义特征“功能”“效应”“功效”以及非语义特征“申请机构”，通过“功能 + 效应”信号组合锁定潜在弱信号，并通过申请机构、申请时间、功效进一步识别新兴技术弱信号，最后应用该方法识别出了“隐形眼镜消毒”领域的潜在新兴技术。

事实上，现有研究较为缺乏对技术发展内在动力与其发展趋势之间关系的解读[47]。而技术发展的内在动力主要涉及市场需求方面，包含问题解决、功能实现等表征因素，这些数据可以通过专家意见获取，但也可以通过结构语义分析方法从专利文献甚至科技评论中自动获取，这可以大幅度地降低主观性、节省人力物力，更好地识别技术未来发展潜力，因此在当前可供使用的数据源及其中可挖掘的信息不断增多的背景下，基于结构语义分析的新兴技术识别方法具有较大的研究价值。

3 结语

自宾夕法尼亚大学沃顿商学院 2000 年出版著作 Wharton on Managing Emerging Technologies 后，新兴技术识别一直都是学者们关注的热点[1]。目前的新兴技术识别研究整体呈现如下特征：

（1）新兴技术识别越来越受到各领域、各层面机构的关注和重视，并从各自目的出发组织开展了诸多的新兴技术识别活动。在国家层面进行的新兴技术识别其目的在于加强国家科技战略规划和整体科研布局的科学性，在企业层面进行的新兴技术识别其目的在于确定企业技术研发重点及投资方向，研究机构进行的新兴技术识别其目的在于寻找未来科研方向和支持做好科研的前瞻布局。新兴技术识别在相关科技战略决策中日益发挥着重要的支撑作用。

（2）不同层面的研究主体识别新兴技术的目的有所不同，相应的，其所采取的方法也有所差异，即不同的新兴技术识别方法有不同的应用场景。

（3）近年来，新兴技术识别预测方法的研究热度不断上升，其中，文献计量及专利分析是应用最为广泛的方法，而机器学习、文本挖掘等方法在其中的应用近几年发展尤为迅速，深度学习方法也开始在新兴技术识别预测领域崭露头角。

总之，近年来新兴技术识别领域在指标体系设计、模型构建、算法运用等方面都产生了较多的创新性成果，但是目前该领域的研究依然存在一些不足，具体如下：

（1）缺乏对新兴技术本质特征的研究，导致研究

针对性不足。目前的研究多的是探索引进和应用一些新的技术、方法,以期不断改进新兴技术发现与识别的效果,但由于对新兴技术本质特征的研究不足,因而常常导致方法层面的改进效果总是欠佳。

(2)具体识别方法对新兴技术识别的解释意义不强。目前的新兴技术识别方法中除了个别特征性指标含义较为明确,其他很多指标都缺乏与新兴技术识别的关联解释,而抽象的数学模型等在这方面的问题则更为突出。

(3)数据时效性制约问题还没有得到有效解决。目前,在新兴技术识别领域的研究更多停留在"回顾性研究"层面,这与数据源及方法的选取两方面因素都有关联。在数据源方面,虽然当前可用的数据源更为多样,但是学者们在实际研究中基本只选取了论文及专利数据,而论文及专利数据的时滞性却是难以回避的问题。

(4)缺乏对多源数据有效融合方法的探索。虽然有很多学者提到,通过多源数据融合计算识别新兴技术可能会有更好的识别效果,但是目前在这方面进行深入探索的学者相对较少,原因在于多源数据的有效融合难度较大,且缺乏专业理论指导和机理性研究成果的支撑。

针对目前新兴技术识别研究领域的不足,未来研究应该在以下方面进行加强:

(1)加强对新兴技术本质特征的研究,构建更为完善的新兴技术识别标准体系,使得新兴技术的识别更具针对性及说服力。

(2)加强具体识别方法对新兴技术的解释性,更好地搭建"识别方法－结果解读"的桥梁。

(3)努力克服数据时效性的制约。在数据层面,进一步挖掘其他数据源的价值和可用性,充分利用当前数据处理方法的高效性及便捷性,发挥各类数据的优势,提高新兴技术识别的效率与准确度;在方法层面,通过优化指标选取尽量地弱化专利等数据的时滞性问题,以更好得强化分析结果的决策支撑价值。

(4)加强对多源数据融合计算的理论与方法的深入探索,从而为新兴技术识别甚至整个技术预测领域开启新的通道。

参考文献:

[1] 戴,休梅克. 沃顿论新兴技术管理[M]. 石莹,译. 北京:华夏出版社,2002.

[2] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology? [J]. Research policy, 2015,44(10):1827－1843.

[3] PORTER A L, ASHTON W B, CLAR G, et al. Technology futures analysis: toward integration of the field and new methods[J]. Technological forecasting and social change,2004, 71(3):287－303.

[4] 黄鲁成,卢文光. 基于属性综合评价系统的新兴技术识别研究[J]. 科研管理,2009, 30(4):190－194.

[5] 任智军,乔晓东,张江涛. 新兴技术发现模型研究[J]. 现代图书情报技术,2016,32(7):60－69.

[6] KYEBAMBE M N, CHENG G, HUANG Y, et al. Forecasting emerging technologies: a supervised learning approach through patent analysis [J]. Technological forecasting and social change, 2017,125:236－244.

[7] SONG K, KIM K, LEE S. Identifying promising technologies using patents: a retrospective feature analysis and a prospective needs analysis on outlier patents[J]. Technological forecasting and social change,2018,128:118－132.

[8] WANG Q. A bibliometric model for identifying emerging research topics[J]. Journal of the association for information science and technology,2018,69(2):290－304.

[9] MUND C, NEUHAUSLER P. Towards an early-stage identification of emerging topics in science -The usability of bibliometric characteristics[J]. Journal of informetrics,2015, 9(4):1018－1033.

[10] SCHIEBEL E, HOERLESBERGER M, ROCHE I, et al. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices[J]. Scientometrics,2010, 83(3): 765－781.

[11] FUJITA K, KAJIKAWA Y , MORI J, et al. Detecting research fronts using different types of weighted citation networks[J]. Journal of engineering and technology management, 2014,32(SI):129－146.

[12] SMALL H, BOYACK K W, KLAVANS R. Identifying emerging topics in science and technology[J]. Research policy, 2014, 43(8):1450－1467.

[13] BREITZMAR A, THOMAS P. The emerging clusters model: a tool for identifying emerging technologies across multiple patent systems [J]. Research policy,2015,44(1):195－205.

[14] FURUKAWA T, MORI K, ARINO K, et al. Identifying the evolutionary process of emerging technologies: a chronological network analysis of World Wide Web conference sessions[J]. Technological forecasting and social change,2015,91:280－294.

[15] 王燕鹏,韩涛,赵亚娟,等. 人工智能领域关键技术挖掘分析[J/OL]. 世界科技研究与发展,2019:1－12[2019－09－05]. https://doi.org/10.16507/j.issn.1006-6055.2019.08.001.

[16] KLEINBERG J. Bursty and hierarchical structure in streams[J]. Data mining and knowledge discovery,2003,7(4):373－397.

[17] CHEN C M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of

the American society for information science and technology, 2006, 57(3): 359 – 377.
[18] GUO H, WEINGART S, BOERNER K. Mixed-indicators model for identifying emerging research areas[J]. Scientometrics, 2011, 89(1): 421 – 435.
[19] PILKINGTON A. Technology portfolio alignment commercialisation: an investigation of fuel cell patenting[J]. Technovation, 2004, 24(10): 761 – 771.
[20] LEE C, KWON O, KIM M, et al. Early identification of emerging technologies: A machine learning approach using multiple patent indicators[J]. Technological forecasting and social change, 2018, 127: 291 – 303.
[21] 刘同,真溱,汤珊红. 利用专利分类,识别和分析新兴技术[J]. 情报理论与实践,2015,38(6):145.
[22] LEE P C, SU H N, Wu F S. Quantitative mapping of patented technology -The case of electrical conducting polymer nanocomposite[J]. Technological forecasting and social change, 2010, 77(3): 466 – 478.
[23] CHO T S, SHIH H Y. Patent citation network analysis of core and emerging technologies in Taiwan: 1997 – 2008[J]. Scientometrics, 2011, 89(3): 795 – 811.
[24] 李瑞茜,陈向东. 基于专利共类的关键技术识别及技术发展模式研究[J]. 情报学报,2018,37(5):49 – 56.
[25] 黄璐,朱一鹤,张嶷. 基于加权网络链路预测的新兴技术主题识别研究[J]. 情报学报,2019,38(4):335 – 341.
[26] LEE S, YOON B, PARK Y. An approach to discovering new technology opportunities: Keyword-based patent map approach[J]. Technovation, 2009, 29(6 – 7): 481 – 497.
[27] DE RASSENFOSSE G, DERNIS H, GUELLEC D, et al. The worldwide count of priority patents: A new indicator of inventive activity[J]. Research policy, 2013, 42(3): 720 – 737.
[28] ERDI P, MAKOVI K, SOMOGYVARI Z, et al. Prediction of emerging technologies based on analysis of the US patent citation network[J]. Scientometrics, 2013, 95(1): 225 – 242.
[29] 李蓓,陈向东. 基于专利引用耦合聚类的纳米领域新兴技术识别[J]. 情报杂志,2015,34(5):35 – 40.
[30] KIM G, BAE J. A novel approach to forecast promising technology through patent analysis[J]. Technological forecasting and social change, 2017, 117: 228 – 237.
[31] JARVENPAA H M, MAKINEN S J, SEPPANEN M. Patent and publishing activity sequence over a technology's life cycle[J]. Technological forecasting and social change, 2011, 78(2): 283 – 293.
[32] 王凌燕,方曙,季培培. 利用专利文献识别新兴技术主题的技术框架研究[J]. 图书情报工作,2011,55(18):74 – 78 + 23.
[33] OGAWA T, KAJIKAWA Y. Assessing the industrial opportunity of academic research with patent relatedness: a case study on polymer electrolyte fuel cells[J]. Technological forecasting and social Change, 2015, 90: 469 – 475.
[34] 白如江,冷伏海,廖君华. 科学研究前沿探测主要方法比较与发展趋势研究[J]. 情报理论与实践,2017,40(5):33 – 38.
[35] 王兴旺,董珏,余婷婷. 基于三螺旋理论的新兴产业技术预测方法探索[J]. 科技管理研究,2019,39(6):108 – 113.
[36] 徐路路,王效岳,白如江. 基于 PLDA 模型与多数据源融合相关性分析的新兴主题探测研究——以石墨烯领域为例[J]. 情报理论与实践,2018,41(4):63 – 69,43.
[37] PAPADIMITRIOU C H, RAGHAVAN P, TAMAKI H, et al. Latent semantic indexing: A probabilistic analysis[J]. Journal of computer and system sciences, 2000, 61(2): 217 – 235.
[38] 董放,刘宇飞,周源. 基于 LDA-SVM 论文摘要多分类新兴技术预测[J]. 情报杂志, 2017,36(7): 40 – 45,133.
[39] CHOI D, SONG B. Exploring technological trends in logistics: topic modeling-based patent analysis[J]. Sustainability, 2018, 10(8): 1 – 26 .
[40] 周源,刘宇飞,薛澜. 一种基于机器学习的新兴技术识别方法:以机器人技术为例[J]. 情报学报,2018,37(9):83 – 99.
[41] ZHOU Y, LIN H, LIU Y F, et al. A novel method to identify emerging technologies using a semi-supervised topic clustering model: a case of 3D printing industry[J]. Scientometrics, 2019, 120(1): 167 – 185.
[42] 刘宇飞,尹力,张凯,等. 基于深度迁移学习的技术术语识别——以数控系统领域为例[J/OL]. 情报杂志:1 – 8[2019 – 09 – 05]. http://kns. cnki. net/kcms/detail/61. 1167. G3. 20190826. 1717. 022. html.
[43] YOON J, CHOI S, KIM K. Invention property-function network analysis of patents: a case of silicon-based thin film solar cells[J]. Scientometrics, 2011, 86(3): 687 – 703.
[44] 李欣,王静静,杨梓,等. 基于 SAO 结构语义分析的新兴技术识别研究[J]. 情报杂志,2016, 35(3):80 – 84.
[45] YANG C, HUANG C, SU J. An improved SAO network-based method for technology trend analysis: A case study of graphene[J]. Journal of informetrics, 2018, 12(1): 271 – 286.
[46] 翟东升,夏军,张杰,等. 基于专利新兴技术弱信号识别方法研究[J]. 情报杂志,2015,34(8): 31 – 36.
[47] XU Y S, HUA X F. Mapping technological trajectories as patent citation networks: Taking the aero-engine industry as an example[C]//KOCAOGLU D F, ANDERSON T R, DAIM TU, et al. 2014 Portland international conference on management of engineering & technology. New York: IEEE, 2014: 2827 – 2835.

作者贡献说明:

王玏:提出论文框架,负责文献数据采集、分析以及论文撰写;

吴新年:提出论文选题,负责论文修改与完善。

王玏，吴新年. 新兴技术识别方法研究综述[J]. 图书情报工作,2020,64(4):125 - 135.

Research on Identification Methods of Emerging Technologies

Wang Le[1,2,3] Wu Xinnian[1,2,3]

[1] Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000

[2] Lanzhou Library of Chinese Academy of Sciences, Lanzhou 730000

[3] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100049

Abstract: [**Purpose/significance**] The paper systematically analyzes the identification methods of emerging technologies at home and abroad, then summarizes research status and analyzes existing problems to provide reference for future research on identification methods of emerging technologies. [**Method/process**] On the basis of literature research and bibliometric analysis, this paper summarized the current mainstream identification methods of emerging technologies and analyzed its characteristics and its impact on the research of identification methods of emerging technologies combined with typical cases, then proposed suggestions for future research on identification methods of emerging technologies. [**Result/conclusion**] The current identification methods of emerging technologies conclude quantitative methods and qualitative methods. The common identification methods of emerging technologies are mainly based on measurement models, bibliometrics and text mining methods, and they are increasingly inclined to be multi-method fusion. However, there are still some deficiencies in the research. Future research on identification methods of emerging technologies should strengthen the exploration of the essential characteristics of emerging technologies, and strengthen the interpretation of specific identification methods to emerging technologies, and promote the timeliness of identification of emerging technologies from data sources and identification methods as well as Strengthen the exploration of effective fusion methods and application methods of multi-source data in the identification of emerging technologies.

Keywords: emerging technology technology identification identification methods